Title: "We only talk about the morning because that's what you ask us": Qualitative development and cognitive testing of candidate items for inclusion in a new rheumatoid arthritis stiffness patient-reported outcome measure

Short title: Development and cognitive testing of candidate items for a new RA stiffness PROM

Authors: Serena Halls MSc PhD[1], Emma Dures MSc PhD CPsychol[1], John R Kirwan MD FRCP FHEA[2], Jon Pollock PhD[1], Gill Baker[3], Avis Edmunds[3], Sarah Hewlett FRCN PhD MA RN[1]

Affiliations: [1]Faculty of Health and Applied Sciences, University of the West of England, Bristol, UK

[2]School of Clinical Sciences, University of Bristol, Bristol, UK

[3]Rheumatology Department, Bristol Royal Infirmary, Bristol, UK

Correspondence to: Serena Halls, Academic Rheumatology Unit, The Courtyard, Bristol Royal Infirmary, Bristol BS2 8HW, UK, Serena.halls@uwe.ac.uk, 0117 342 4972.

Abstract

Objective: To qualitatively develop and test a set of candidate items for a new RA stiffness patient-reported outcome measure (PROM) that capture the patient perspective. This is an essential first step in PROM development, prior to quantitative development, assessment and validation.

Methods: Focus groups further examined the previously developed stiffness conceptual model and explored the patient perspective regarding stiffness assessment. Data were analysed using thematic analysis. An iterative process of item development was then performed by the expert study team of researchers, patients and clinicians, based on the two qualitative datasets and informed by measurement theory and guidelines. Finally, these candidate items were tested using formal cognitive interview methodology and subsequently refined.

Results: Sixteen RA patients from the UK participated in focus groups. Data confirmed the conceptual model of the RA patient experience of stiffness and provided insight into stiffness assessment including suggestions regarding patient-relevant stiffness assessment categories such as impact, location, and timing. These data informed the development of 77 candidate stiffness PROM items, including multiple formats for some. Eleven RA patients participated in cognitive interviews. Minor changes were made to items to enhance understanding and 32 items were removed, resulting in 45 candidate PROM items.

Conclusion: Rigorous qualitative methodology and considerable patient involvement has underpinned items for a new RA stiffness PROM with strong content validity. Crucially, patient involvement broadened assessment beyond ‘early morning stiffness duration,’ which may address existing PROM limitations. Items are now suitable for quantitative item reduction, structural development of the final PROM, and validation.

Key words: Rheumatoid arthritis, stiffness, morning stiffness, early morning stiffness, patient-reported outcome measure

Funding: University of the West of England

Introduction

RA is a chronic, systemic, inflammatory condition causing synovitis and resulting in pain, swelling and stiffness [1]. Early clinical observations recognised morning stiffness (MS) or early morning stiffness (EMS) as a common feature of RA [2] and as such its subjective assessment was included in original classification [3] and remission [4] criteria, and composite assessments of disease activity (for example, [5]). However, EMS/MS was omitted from remission and classification criteria updates, the DAS, and the RA core set because of poor measurement properties of available patient-reported outcome measures (PROMs) [6-10]. Therefore, there is currently no obligation to assess stiffness in routine clinical care or clinical trials.

Despite this, stiffness remains frequently queried in research and clinical contexts and is a relevant symptom to RA patients. In research, MS is regularly employed as a study inclusion criterion and as an outcome measure [11]. Specifically, MS is one of the few PROs (alongside pain, function and patient global assessment (PtG)) reported in ≥25% of research studies [12]. It has also been employed as the primary outcome measure in current research into timed-release glucocorticoid treatments [13], results from which will inform patient treatment in clinical practice. Clinically, stiffness is commonly observed in an RA population with a recent prevalence estimate of between 70-75% regardless of treatment status [14]. Clinicians report awareness [15] and regular assessment of MS [16], and use it as an important variable in decision-making for changing patients' medication [17]. Presence of MS has also been found to be a determinant of earlier initiation of DMARDs in an RA population [18]. Importantly, patients report that stiffness has a significant impact on their life [15], specifically work life [19]. Yet despite the relevance and common use of stiffness, until recently it has rarely been the focus of research. However, recent qualitative work furthered understanding of the patient experience of this symptom and developed a conceptual model capturing patients' experiences of RA stiffness [20]. This was reinforced by work performed in a US patient sample [21].

In work into remission and flare, stiffness assessment has been identified as an area requiring investigation [22, 23]. This is particularly relevant to determining remission in a clinical context as it has been suggested that MS is likely to affect patients being unable to meet the stringent ACR/EULAR Boolean remission criteria [24]. Yet both a systematic literature review of stiffness PROMs in the assessment of low disease activity or remission [25] and an update of that review across all disease activity states [26] concluded that there was insufficient evidence supporting current stiffness PROMs, no indication of which stiffness PROM to use, and that no current PROMs had been developed with patient involvement. This is a fundamental concern as PROM development guidelines state that the purpose of PROMs are to capture what is relevant to the population of interest [27]. The importance of appropriate content validity in the development of PROMs has also been reinforced by the recommendation to include qualitative underpinning and involve the relevant population in PROM development [28, 29]. A lack of content validity in current stiffness PROMs was indicated in recent qualitative work where patients described stiffness as more than just severity and duration and not exclusively experienced in the morning [20, 21]. This contrasts with the most commonly used PROM assessments, which capture only morning stiffness severity or duration [25]. Given the apparent inconsistency between current stiffness assessment and the patient experience of this symptom, the aim of this study was to: 1) clarify and/or expand the proposed stiffness conceptual model and obtain patient views on assessment; 2) develop and then 3) qualitatively test a set of candidate items for a new RA stiffness PROM that captures the patient perspective and demonstrates appropriate content validity resulting from development in accordance with best practice guidelines [27-29].

Patients and Methods

Study methods involved patient focus groups, item development and patient cognitive interviews. Ethics approval was granted by the Leeds East Research Ethics Committee (13YH0050).

Focus groups

Patients with clinician-diagnosed RA [3, 8], aged ≥18 years, with the ability to speak English unaided were recruited from a rheumatology outpatient department in a UK hospital. Patients were invited to participate by a researcher (SH) or research nurse. Purposeful sampling was employed, and a sampling frame ensured patients with a range of age, sex and disease duration were recruited. Sample size was determined based on current recommendations [30]. Prior to each focus group, patients gave written informed consent and completed a brief questionnaire containing demographic and clinical information (Health Assessment Questionnaire (HAQ) [31]; PtG [32]).

A focus group topic guide was developed: Part A asked about participants experience of stiffness to confirm or elaborate on the stiffness conceptual model developed in previous qualitative work [20]; Part B asked participants' about their views regarding how stiffness should be assessed. An iterative process which allowed ideas and concepts identified in early data collection to be explored in subsequent data collection was employed [33]. Focus groups were moderated by two researchers (SH, ED) and audio-recorded, transcribed verbatim and managed using Nvivo 10 [34], Microsoft Office Word and Excel 2013. Deductive analysis is driven by theoretical influences such as existing theory, previous research, or coding frames [35]. The coding frame identified in previous stiffness qualitative work [20] (Part A), and a framework developed based on questionnaire design literature [36] including the broad categories of: stem questions and anchors, response options, timeframe, layout and format (Part B) were applied to these data.

Item (question) development

Item development was then performed, involving an iterative process of discussion with the expert study team and subsequent item refinement. Items were developed based on the combined qualitative data from this focus group study and from previous qualitative work [20]. The process was informed by PROM development guidelines [27-29]. It was also influenced by consideration of measurement theory including the need for measurement tools to

demonstrate appropriate measurement properties [37], and consideration of the 2014 OMERACT Filter which evaluates outcome measures against the concepts of truth (is the measure unbiased?), discrimination (is the measure sensitive and reliable?) and feasibility (is the measure understandable and time efficient for ease of use in clinical and research environments?) [38].

Cognitive interviews

All candidate items were then reviewed by each patient during cognitive interviews. Cognitive interviewing is a formal research methodology capturing the cognitive process of item response enabling identification of difficulties with questions or response option interpretation [39]. A separate sample of patients were recruited using the same criteria and methods as described above. Cognitive interviews were performed by one researcher (SH). Patients were asked to complete the candidate items as they would any questionnaire but to explain what they were thinking as they read the question and judged their answers. Data were audio-recorded, transcribed verbatim and managed using Nvivo 10 [34] Microsoft Office Word and Excel 2013. Data were analysed deductively based on the four headings of the four-stage cognitive model: 'Understanding' (did the patient understand the question?), 'Retrieval' (was the patient able to retrieve from memory the information required?), 'Judgement' (was the patient able to make a judgement?) and 'Response' (was the patient able to select an appropriate response?) [39]. The candidate items were then refined based on the cognitive interview data and discussion with the expert study team to develop the final list of candidate items for the new RA stiffness PROM.

Results

Overall, 27 RA patients with a range of demographics participated (Table 1).

Focus groups

16 RA patients (11 female (68.8%), median age 64.5 years (interquartile range (IQR) 57-72 years, median disease duration 6.5 years (IQR 3.5-13 years) participated across three focus groups (Table 1), each lasting approximately 120 minutes.

Focus group data supported our previous qualitative work [20] in a new sample of patients using a different method of data collection. Each previously identified theme within the conceptual model of the RA patient experience of stiffness (*Part of having RA; Linked to behaviour and environment; Local and widespread; Highly variable; Impacts on daily life; Requires self-management*) [20] was supported by the focus group data, enhancing the robustness of the model (Table 2). Patients are identified by study number, gender, and study identifier, for example: Study 2, Male, study identifier 10 [2-M-10].

Investigation into the patient perspective of stiffness assessment provided insight into stem question categories and their relevance to patients, and patient preferences regarding response options, timeframe, and format. Participants raised stem question categories including impact, location, timing, and stiffness after immobility (Table 3). Other considerations for measurement that patients raised included the relationship between stiffness and symptoms such as pain and inflammation: *"[…] stiffness, yes or no, with pain, yes or no, with swelling, yes or no"* [2-M-10], and the individual nature of stiffness: *"It's how you feel, not the average or somebody else"* [2-M-05]. Patients provided clarification that 'stiffness' was a patient relevant word*: "Well, stiffness would be used, wouldn't it?" […] Everybody says it, don't they?"* [2-M-10] *"Yes"* [multiple responders].

Importantly, patients reported that they found the concept of stiffness duration difficult to answer, feeling it was hard to remember or quantify: *"[…] when you come to the doctors and they say how long does it last, well, it's about that long but it's a guess really"* [2-F-09] *"Yes and you suddenly realise you haven't got it then"* [2-F-08] [agreement]. Patients stated that they were unsure what the start or endpoints for the clinician's questions on duration were: *"[…] we're not working in the same way that the doctors are working, you know. In our minds we're not sort of sitting there timing it"* [2-F-16] [agreement and laughter] *"Oh I am thoroughly unstiffened, no […] that's not the real world"* [2-F-16]. Additionally, duration questions were

not felt to capture the whole experience of stiffness because they focused on morning stiffness: *"[…] if they're [clinician] just looking at morning stiffness, then that doesn't capture the general on-going seizing up through the day stiffness, sometimes it does but quite often, well it doesn't at all and morning stiffness for me is mostly where my RA is in a flare or it's not well managed […] at the moment it's sort of fairly okay-ly managed so I'm not getting a lot of morning stiffness, but I do seize up through the day"* [2-F-16]. Patients reflected that it was the clinician's insistence on asking about morning stiffness that led to it being discussed, but limiting it to mornings was not particularly relevant to them: *"I think, now we're talking about it, we only talk about the morning because that's what you ask us"* [2-F-01] *[laughter] "That's right, that's right" [2-M-05] "But it is, it's after anytime"* [2-F-01] *"Any time of day really"* [2-F-04].

Patients also discussed their preferences regarding relevant response options and layout focusing on the importance of simplicity and brevity: *"Yes, less options"* [2-M-06] *"Yes, less options"* [2-M-10] *"More straightforward questions, less options"* [2-M-06].

Item development

These focus group data on stiffness experience and assessment were combined with data from the previous qualitative work underpinning the conceptual model of the patient experience of stiffness [20] and were mapped on to each other to form the basis for item development. From this, stem questions, response options, timeframe, and layout and format were designed in an iterative process involving the expert study team of researchers, patients, and clinicians. This process was informed by measurement theory [37, 38], PROM development guidelines [27-29], and consideration of the purpose of the PROM. Iterations of item development were captured in tracking tables [27] to enable checking and moving backward and forward through these data.

Specifically in relation to the development of item response options, measurement theory literature suggested that four or five response options are preferable given that they place less burden on responders yet are still precise [40], while a number of rheumatology PROMs with

which RA patients may be familiar (e.g. HAQ [31] and Bristol Rheumatoid Arthritis Fatigue Multidimensional Questionnaire (BRAF) [41]) use four response options. This was consistent with the focus group preference for few and simple response options; thus both the literature and research informed the development of item response options. Additionally, 24 items were presented in several formats (visual analogue and numerical rating scales (VAS/NRS) and Likert scales), in an attempt to discern the preferred format.

Overall, a bank of 77 candidate items were designed. These included the 24 items presented in multiple formats and six current stiffness PROM items (Table 4) identified in the literature review [25, 26]. Items were formatted into a paper questionnaire for further qualitative testing using cognitive interviews. To reduce participant burden, only one version of items presented in multiple formats were included in the questionnaire but all versions were discussed with all participants.

Cognitive interviews

11 RA patients (7 female (63.6%), median age 61 years (IQR 56-77 years, median disease duration 5 years (IQR 2-20 years, Table 1) participated in cognitive debriefing of all candidate items, each lasting approximately 60 minutes.

Items were generally well understood. No difficulties were identified under the heading of 'Retrieval'. Minor difficulties were identified with retained items under the headings of 'Understanding' (n=34), 'Judgement' (n=13), and 'Response' (n=18). These required changes, which although only minor, significantly improved the clarity and understanding of items. An example of a change made under the heading of 'Understanding' was in the item 'Has stiffness made it difficult to do fine movements? For example, do up buttons on a shirt or cardigan?' The seasonal nature of the item was highlighted: *"Well, it's this time of year, you don't do up buttons, do you? [3-M-03]"*. This item was therefore edited to be more broadly relevant, based on an example suggested in the previous qualitative work: 'Has RA stiffness made it difficult to do fine movements (for example, write with a pen)?' An example under the heading of 'Judgement' included the observation that items requiring factual information were difficult to

make a judgement on. For example with the item 'How much has joint damage contributed to your experience of stiffness?' (intended to differentiate between inflammatory and mechanical stiffness), one patient stated: *"That one is a bit hard to know because unless you have an x-ray or something you don't really know"* [3-F-09]. These items were either edited or removed. Under the heading of 'Response', it appeared that some layouts led participants to select answers they did not intend: *"Oh, did you say that one [answer] was 'very much so'?" [SH] […] "Oh yes, got it in the wrong one, haven't I?"* [3-F-06]. As a result, the layout was changed, substituting a grid for boxes to ensure that the appropriate response option could be marked. Also in relation to 'Response', items using different response option formats (VAS/NRS/Likert scale) were discussed. Although patients described advantages and limitations with all formats, NRS were described as easy and clear: *"I think if you grade it up to 10 it is probably easier in people's minds […]"* [3-F-09], and therefore were retained.

Overall, minor changes were made to items for consistency and to improve clarity and understanding. Thirty-six of the 77 items were removed either because they had alternative formats to the retained NRS format, or because they contained information already captured in other items; four items were added based on patient suggestions. These included items to capture stiffness location, stiffness perceived as being different to usual and two items capturing stiffness perceived to be a result of joint damage. The current stiffness PROM item capturing MS on a VAS was also replaced by another current stiffness PROM item from the literature capturing stiffness rather than MS (Table 4). A final bank of 45 candidate items (Table 5), including six current stiffness PROM items from the literature (Table 4), were defined as suitable to take forward into further testing for quantitative item reduction, structural development of the final PROM and validation.

Discussion

This body of work demonstrates that collaboration with patients to better characterise and assess the patient experience of stiffness, results in a candidate list of items that assess

stiffness in an understandable and patient relevant way. Previous qualitative work in the UK [20] and US [21] demonstrated considerable similarities and improved understanding of the patient experience of stiffness in patients with RA. The conceptual model of RA stiffness [20] which emerged from that work was further explored here as part of the focus group discussions, where the model was confirmed in a new population of RA patients using a different qualitative method, enhancing the rigor of the model [42]. This is key as it supports universality of the conceptual model themes with both local and international work.

Importantly, these qualitative studies demonstrated that RA patient descriptions and assessment of stiffness are much broader than, and in some cases inconsistent with, those captured in existing stiffness PROMs. While this had been suggested in previous qualitative work [20, 21], the results presented here enabled confirmation and further elaboration of these suggestions in an approach that focused on the development of appropriate stiffness assessment. For example, while current stiffness PROMs capture morning stiffness duration or severity [25], RA patients describe stiffness as not exclusive to the morning [20, 21] and in the current study expressed difficulties with responding to items that focus solely on this timeframe. Furthermore, the apparent multidimensional nature of stiffness identified in the conceptual model particularly the impact that stiffness has on patients' lives [20] was here identified as a relevant stem question category (Table 3), and challenges the narrow focus of current PROM items focusing on severity and duration alone.

The relevance of the concept of impact in patients with chronic conditions is not unique to the symptom of stiffness. The impact triad is a concept developed by patients and researchers, who propose that the severity of an outcome, its importance to patients, and their ability to self-manage it, all combine to form impact, and that these should all be captured in patient-reported outcome assessment [43]. Impact is also a key component of other recently developed and well validated rheumatology outcome measures including the Psoriatic Arthritis Impact of Disease (PsAID) [44].

The biomedical interpretation of stiffness relates the symptom to the circadian rhythms of pro-inflammatory cytokines such as IL-6 which increase in the early morning [45]. However, patient

descriptions of stiffness not being exclusively experienced in the morning period challenge this biomedical interpretation in relation to patient-reported assessment of stiffness. This finding is supported by work with patients with PMR who similarly reported that stiffness is not purely experienced in the morning [46] and a recent Delphi study in the development of a core domain set for PMR where patients expressed a preference for 'stiffness' rather than 'morning stiffness' [47]. Work in PMR has also questioned the adequacy of duration as part of stiffness assessment [46]. This was consistent with the patient dislike of duration expressed in this study, and is concerning when considering that stiffness duration items are most frequently implemented in the assessment of stiffness in research trials [12].

It is important to reiterate here that the purpose of a PROM is to capture the patient experience [27]. No current stiffness PROM appears to have been developed with such substantial involvement from the target population [26]. This demonstrates the added value of the new PROM items compared to those currently available. Poor content validity of current stiffness items may explain the inadequate psychometric performance of these items reported in the literature [6-10]. It also may explain the challenges reported with current items including that patients find completing duration items difficult, are often forced to report a cut-off time [48], and that patients have sometimes reported that they have no stiffness in a leader question asking about stiffness duration, only to later quantify the severity of that non-existent stiffness in a follow-up question [49]. This further supports the need for the involvement of the relevant population in the development of PROMs to enhance content validity and ensure that the patient experience is captured in outcome assessment.

These results also relate to work within the OMERACT group. This article has addressed some of the key areas discussed at the breakout group within the OMERACT 2014 RA Flare Group Workshop [50] including the dislike of stiffness duration as a measure and the importance of the impact of stiffness from the patient perspective. It also drives forward the importance of further research on the topic of stiffness as highlighted on the research agenda of the report from the inaugural stiffness special interest group at OMERACT 2016 [26].

This study provides evidence for the importance of performing cognitive interviews during PROM development, as suggested in guidelines [27-29]. Whilst the cognitive interviews only identified minor difficulties with items, the changes subsequently made to items were crucial to enhance acceptability to the intended population. If these difficulties had not been identified then subsequent inaccurate data collection may have resulted at later stages.

These studies only recruited English-speaking patients from one UK based rheumatology outpatient department, which may affect generalisability. However, conceptual similarities within qualitative work in a heterogeneous US population have been reported [21], suggesting that these findings may be relevant in such populations. A further limitation relates to cognitive interviewing which is performed in a controlled research environment that may differ from use in an applied research or clinical setting. However, it is suggested that cognitive interviewing will identify the most significant problems with items [39], thus ensuring that items are appropriate for use in applied settings. Item development was based on data generated in qualitative studies with small samples. However, qualitative studies were rigorously performed in an iterative process with clinical, patient and research experts and data saturation [30] was achieved in all studies. Furthermore, the qualitative nature of the item development is a key strength of this work, consistent with recommended PROM development methodology [27-29] and recommendations from EULAR and OMERACT.

The final 45 candidate items have face and content validity. They are now suitable for quantitative research to establish the smallest yet most internally consistent group of items to form a new RA stiffness PROM. This can then be subject to psychometric property evaluation, including construct and criterion validity, reliability, sensitivity to change, floor and ceiling effects [37]. After administration of the items to a wider patient sample (e.g. socio-demographic, disease activity), classical and modern psychometric approaches can then be applied to determine the dimensionality of the construct, identify item redundancies and define a final, optimal set of items that best measures patient-reported stiffness. This work recognises stiffness as a relevant and recordable patient symptom and is a significant step towards a standardised assessment tool with appropriate measurement properties.

Key messages

- RA patients describe and assess stiffness using wider concepts than captured in current stiffness PROMs
- The RA patient experience of stiffness informed the development of items for a new PROM
- This is a key step towards standardised stiffness assessment and recognition of a relevant patient symptom

Acknowledgements

The authors would like to thank the patients who participated in this research and the clinical teams who facilitated recruitment.

Conflict of interest

JK has received support from Horizon Pharma Inc. to attend scientific meetings. ED and SH hold an unrestricted educational grant from Pfizer. All other authors have declared no conflicts of interest.

Funding

This work was supported by a PhD studentship funded by the University of the West of England, Bristol, UK (grant number: 430B180).

References

1 Hill J. *Rheumatology nursing. A creative approach.* Chichester: John Wiley & Sons Ltd; 2006.
2 Scott JT. Morning stiffness in rheumatoid arthritis. Ann Rheum Dis 1960;19:361-8.
3 Arnett FC, Edworthy SM, Bloch DA, et al. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. Arthritis Rheum 1988;31(3):315-24.
4 Pinals RS, Masi AT, Larsen RA. Preliminary criteria for clinical remission in rheumatoid arthritis. Arthritis Rheum 1981;24(10):1308-15.
5 Lansbury J. Report of a three-year study on the systemic and articular indexes in rheumatoid arthritis; theoretic and clinical considerations. Arthritis Rheum 1958;1(6):505-22.
6 van der Heijde DM, van 't Hof MA, van Riel PL, et al. Judging disease activity in clinical practice in rheumatoid arthritis: first step in the development of a disease activity score. Ann Rheum Dis 1990;49(11):916-20.
7 Felson DT, Anderson JJ, Boers M, et al. The American-College-of-Rheumatology Preliminary Core Set of Disease-Activity Measures for Rheumatoid-Arthritis Clinical-Trials. Arthritis Rheum 1993;36(6):729-40.
8 Aletaha D, Neogi T, Silman AJ, et al. 2010 Rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. Arthritis Rheum 2010;62(9):2569-81.
9 Felson DT, Smolen JS, Wells G, et al. American College of Rheumatology/European League Against Rheumatism Provisional Definition of Remission in Rheumatoid Arthritis for Clinical Trials. Arthritis Rheum 2011;63(3):573-86.
10 Kay J, Upchurch KS. ACR/EULAR 2010 rheumatoid arthritis classification criteria. Rheumatology (Oxford) 2012;51 Suppl 6:vi5-9.
11 Cutolo M. How should morning function in rheumatoid arthritis be assessed? Bibliographic study of current assessment. Scand J Rheumatol 2011;40:17-22.
12 Kalyoncu U, Dougados M, Daures JP, Gossec L. Reporting of patient-reported outcomes in recent trials in rheumatoid arthritis: a systematic literature review. Ann Rheum Dis 2009;68(2):183-90.
13 Buttgereit F, Mehta D, Kirwan J, et al. Low-dose prednisone chronotherapy for rheumatoid arthritis: a randomised clinical trial (CAPRA-2). Ann Rheum Dis 2013;72(2):204-10.
14 Strand V, Holt RJ, Saunders KC, et al. Prevalence of Morning Stiffness in a US Registry Population of Rheumatoid Arthritis Patients [abstract]. Arthritis Rheumatol 2014;66:S178-S.
15 da Silva JAP, Phillips S, Buttgereit F. Impact of impaired morning function on the lives and well-being of patients with rheumatoid arthritis. Scand J Rheumatol 2011;40:6-11.
16 Hurkmans EJ, Li L, Verhoef J, Vliet Vlieland TP. Physical therapists' management of rheumatoid arthritis: results of a Dutch survey. Musculoskeletal Care 2012;10(3):142-8.
17 Soubrier M, Zerkak D, Gossec L, Ayral X, Roux C, Dougados M. Which variables best predict change in rheumatoid arthritis therapy in daily clinical practice? J Rheumatol 2006;33(7):1243-6.
18 Pappas DA, Kent JD, Greenberg JD, Mason MA, Kremer JM, Holt RJ. Delays in Initiation of Disease-Modifying Therapy in Rheumatoid Arthritis Patients: Data from a US-Based Registry. Rheumatology and Therapy 2015;2(2):153-64.
19 Westhoff G, Buttgereit F, Gromnica-Ihle E, Zink A. Morning stiffness and its influence on early retirement in patients with recent onset rheumatoid arthritis. Rheumatology 2008;47(7):980-4.
20 Halls S, Dures E, Kirwan J, et al. Stiffness is more than just duration and severity: a qualitative exploration in people with rheumatoid arthritis. Rheumatology 2015;54(4):615-22.
21 Orbai AM, Smith KC, Bartlett SJ, De Leon E, Bingham CO. "Stiffness Has Different Meanings, I Think, to Everyone": Examining Stiffness From the Perspective of People Living With Rheumatoid Arthritis. Arthrit Care Res 2014;66(11):1662-72.

22 Bykerk VP, Lie E, Bartlett SJ, et al. Establishing a Core Domain Set to Measure Rheumatoid Arthritis Flares: Report of the OMERACT 11 RA Flare Workshop. J Rheumatol 2014;41(4):799-809.
23 van Tuyl LHD, Hewlett S, Sadlonova M, et al. The patient perspective on remission in rheumatoid arthritis: "You've got limits, but you're back to being you again'. Ann Rheum Dis 2015;74(6):1004-10.
24 Pappas DA, Holt RJ, Shan Y, et al. The Influence of Patient Reported Morning Stiffness on Patient Global Assessment in Rheumatoid Arthritis Patients Not Achieving ACR/EULAR Boolean Remission in a Large US Registry [abstract]. Arthritis Rheumatol. 2015;67(Supplement 10):3202-3.
25 van Tuyl LHD, Lems WF, Boers M. Measurement of stiffness in patients with rheumatoid arthritis in low disease activity or remission: a systematic review. BMC Musculoskel Dis 2014;15.
26 Halls S, Hewlett S, Mackie, S.L, et al. Stiffness Is the Cardinal Symptom of Inflammatory Musculoskeletal Diseases, Yet Still Variably Measured: Report from the OMERACT 2016 Stiffness Special Interest Group. J Rheumatol 2016; online December 15.
27 US Department of Health and Human Services, Food and Drug Administration. Guidance for industry: patient reported outcome measures—use in medical product development to support labeling claims. December 2009. Available freom: www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM193282.pdf.
28 Patrick DL, Burke LB, Gwaltney CJ, et al. Content validity—establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO good research practices task force report: part 1—eliciting concepts for a new PRO instrument. Value in Health 2011;14(8):967-77.
29 Patrick DL, Burke LB, Gwaltney CJ, et al. Content validity—establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO Good Research Practices Task Force report: part 2—assessing respondent understanding. Value in Health 2011;14(8):978-88.
30 Guest G, Bunce A, Johnson L. How many interviews are enough? An experiment with data saturation and variability. Field Method 2006;18(1):59-82.
31 Fries JF, Spitz P, Kraines RG, Holman HR. Measurement of Patient Outcome in Arthritis. Arthritis Rheum 1980;23(2):137-45.
32 Van der heijde DMFM, Vant hof M, Van riel PLCM, Van de putte LBA. Development of a Disease-Activity Score Based on Judgment in Clinical-Practice by Rheumatologists. J Rheumatol 1993;20(3):579-81.
33 Ritchie J, Lewis J. *Qualitative Research Practice: A Guide for Social Science Students and Researchers*. London: Sage Publications Ltd; 2003.
34 Nvivo Qualitative Data Analysis Software (10) [computer programme]. 2012.
35 Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006;3(2):77-101.
36 Streiner DL, Norman GR. *Health measurement scales: a practical guide to their development and use*. Fourth edition ed: Oxford University Press, USA; 2008.
37 Terwee CB, Bot SDM, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 2007;60(1):34-42.
38 Boers M, Kirwan JR, Gossec L, et al. How to Choose Core Outcome Measurement Sets for Clinical Trials: OMERACT 11 Approves Filter 2.0. J Rheumatol 2014;41(5):1025-30.
39 Willis GB. *Cognitive Interviewing. A Tool for Improving Questionnaire Design.* London: Sage Publications Ltd.; 2005.
40 Fries J.F BB, Bjorner J, Rose M. More relevant, precise, and efficient items for assessment of physical function and disability: moving beyond the classic instruments. Ann Rheum Dis 2006;65 (Suppl 3):iii16-iii21.
41 Nicklin J, Cramp F, Kirwan J, Greenwood R, Urban M, Hewlett S. Measuring Fatigue in Rheumatoid Arthritis: A Cross-Sectional Study to Evaluate the Bristol Rheumatoid Arthritis

Fatigue Multi-Dimensional Questionnaire, Visual Analog Scales, and Numerical Rating Scales. Arthrit Care Res 2010;62(11):1559-68.
42 Mays N, Pope C. Rigour and Qualitative Research. Brit Med J 1995;311(6997):109-12.
43 Sanderson TC, Hewlett SE, Flurey C, Dures E, Richards P, Kirwan JR. The Impact Triad (Severity, Importance, Self-management) as a Method of Enhancing Measurement of Personal Life Impact of Rheumatic Diseases. J Rheumatol 2011;38(2):191-4.
44 Gossec L, de Wit M, Kiltz U, et al. A patient-derived and patient-reported outcome measure for assessing psoriatic arthritis: elaboration and preliminary validation of the Psoriatic Arthritis Impact of Disease (PsAID) questionnaire, a 13-country EULAR initiative. Ann Rheum Dis 2014;73(6):1012-9.
45 Straub RH, Cutolo M. Circadian rhythms in rheumatoid arthritis: implications for pathophysiology and therapeutic management. Arthritis Rheum 2007;56(2):399-408.
46 Twohig H, Mitchell C, Mallen C, Adebajo A, Mathers N. "I suddenly felt I'd aged": A qualitative study of patient experiences of polymyalgia rheumatica (PMR). Patient Educ Couns 2015;98(5):645-50.
47 Helliwell T, Brouwer E, Pease CT, et al. Development of a Provisional Core Domain Set for Polymyalgia Rheumatica: Report from the OMERACT 12 Polymyalgia Rheumatica Working Group. J Rheumatol 2016;43(1):182-6.
48 Hazes JMW, Hayton R, Burt J, Silman AJ. Consistency of Morning Stiffness - an Analysis of Diary Data. Brit J Rheumatol 1994;33(6):562-5.
49 Vlieland TPMV, Zwinderman AH, Breedveld FC, Hazes JMW. Measurement of morning stiffness in rheumatoid arthritis clinical trials. J Clin Epidemiol 1997;50(7):757-63.
50 Orbai AM, Halls S, Hewlett S, et al. More than Just Minutes of Stiffness in the Morning: Report from the OMERACT Rheumatoid Arthritis Flare Group Stiffness Breakout Sessions. J Rheumatol 2015;42(11):2182-4.

Tables and Figures

Table 1: Participant clinical and demographic data

	FG 1	FG 2	FG 3	CI	Total
No. of participants	5	5	6	11	27
Age					
Mean (SD)	66 (13.6)	68.4 (7.6)	58.8 (9.4)	65.5 (10.4)	64.6 (11.1)
Median (IQR)	67 (53.5-75)	72 (59.5-75.5)	59.5 (51.3-66.8)	61 (56-77)	64 (56-73)
Sex					
Female	4	3	4	7	18
Male	1	2	2	4	9
Disease duration					
Mean (SD)	5.8 (5.4)	11.2 (7.4)	13.8 (13.7)	9.5 (8.6)	10.1 (9.9)
Median (IQR)	4 (1.5-11)	10 (5.5-17.5)	6 (3.5-29.8)	5 (2-20)	6 (3-16)
HAQ*					
Mean (SD)	1.125 (0.7)	1.65 (0.5)	1.6 (0.8)	1.1 (0.6)	1.3 (0.7)
Median (IQR)	1.375 (0.4-1.8)	1.75 (1.25-2)	1.875 (0.8-2.3)	1.2 (0.5-1.5)	1.5 (0.7-1.9)
PtG					
Mean (SD)	4.12 (3.4)	4.7 (1.9)	3.8 (1.5)	5.0 (2.0)	4.5 (2.3)
Median (IQR)	4.7 (0.3-7.7)	5.0 (2.6-6.7)	4.3 (2.2-5.2)	5.8 (3.3-6.8)	5.0 (2.4-6.1)

FG=Focus group; CI=Cognitive interview; HAQ=Health Assessment Questionnaire (0-3, 3=most disabled); PtG=Patient Global Assessment (0-10, 0=very well, 10=very badly); *=data from two patients missing

Table 2: Patient quotes to support the stiffness conceptual model

Original conceptual model	Example quote supporting original model
Part of having RA	*"I've always thought it [stiffness] was part of what we've got, actually." [2-M-06] [Agreement]*
Linked to behaviour and environment	*"It's either keeping the same position for too long, like sitting on an aeroplane, for example. Or it's when I've exercised too much" [2-F-01] "Yeah, yeah" [2-M-05]*
Local and widespread	*"My hips were stiff, my knees were stiff, everything was really hurting and that was within four hours" [2-F-14]*
Highly variable	*"[…] some days you're going to be more stiff than others" [2-F-09]*
Impacts on daily life	*"Sometimes stiffness is prevention of doing things, like I used to enjoy sewing and threading a needle. I can't do that any longer, and that's the stiffness, it's not the pain. There's no pain in threading a needle, but I just can't do it because my fingers […] parts of me just don't work in the way that they should do […]" [2-F-08]*
Requires self-management	*"[…] sometimes if the stiffness is bad you have to move your joints before you can actually get out of bed." [2-F-16] "That's right." [2-M-11]*

Table 3: Patient quotes to support stem question categories

Stem question categories	Example quote supporting stem question categories
Stem question: Impact	*"I know it's straight forward questions but it's serious questions for people that can't do it […] Comb your hair, brush your teeth, general daily, you think of what you do every day when you get up" [2-M-10]*
Stem question: Location	*"You know the picture they have of a person […] with the massive hands […] I always go to my consultant, that's how it feels [laughs] […] it would be quite nice if you know, you could say, these bits are stiff" [2-F-03]*
Stem question: Timing	*"[…] is there any aspect that particularly springs out when you think about assessing stiffness?" [SH] "When is it worse throughout the day? And is it on waking, is it mid-morning, is it lunchtime, is it afternoon, is it when you feel you're tired? […] I think it's important that you know which parts of the day that individuals have the worst problems? [2-F-08]*
Stem question: Stiffness after immobility	*"So it's not only during the night or first thing in the morning, it's also" [2-F-02] [agreement] "It could be anytime" [2-F-04] "Yes, you're right. Sitting here, for example" [2-M-05] [laughter] "Exactly!! After a period of immobility, whether it's asleep or you're awake" [2-F-01] "That's a good word, I like that […] Immobility. If you've been immobile for, I don't know, an hour. Whatever. Certainly longer. Then, how are your joints then? Nobody has asked that" [2-M-05]*

Table 4: Current stiffness PROM items from literature

Item concept	Item
MS severity	How would you describe the overall level of morning stiffness you have had from the time you wake up? 11-point NRS (0=No stiffness, 10=Very severe stiffness)
MS severity	How would you describe the overall level of morning stiffness you have had from the time you wake up? 100mm VAS (0=No stiffness, 10=Extreme stiffness)
MS severity	How would you describe the overall level of morning stiffness you have had from the time you wake up? 5-option Likert scale (No stiffness, Mild stiffness, Moderate stiffness, Severe stiffness, Very severe stiffness)
MS duration	Were your joints stiff when you woke up today? (Yes/No) If yes, how long did this extra stiffness last? 6-option Likert scale (Less than 30 minutes, 30 minutes to an hour, 1-2 hours, 2-4 hours, More than 4 hours but less than all day, All day)
MS duration	How long does your morning stiffness last from waking until maximum improvement occurs? 3-option Likert scale (Up to 1 hour, 1-3 hours, More than 3 hours)
MS duration	How long does your morning stiffness last from waking until maximum improvement occurs? Minutes/Hours
Severity*	Please circle the number that best describes the severity of your RA stiffness over a usual week when you are not in a flare? 11-point NRS (0=No stiffness, 10=Extreme stiffness)

MS=Morning stiffness; NRS=Numerical rating scale; VAS=Visual analogue scale; *=replaced MS severity item on VAS following cognitive interviews; Item sources provided in [25, 26]

Table 5: Final bank of 45 candidate items for new RA stiffness PROM

Item no.	Item wording	Item response options
1	Do you have any joints that are permanently stuck?	Yes/No
2	Over the past 7 days when have you experienced RA stiffness?	Tick all that apply (In the night, In the morning, In the afternoon, In the evening, None of these)
3	Have you experienced RA stiffness in your joints over the past 7 days?	4-option Likert scale (No, not in any of my joints, Yes, in a few of my joints, Yes in many of my joints, Yes, in all of my joints)
4	Over the past 7 days have you experienced RA stiffness all over?	Yes/No
5	Over the past 7 days has your RA stiffness been different to usual for you?	5-option Likert scale (It has been much better than usual, It has been better than usual, It has been the same as usual, It has been worse than usual, It has been much worse than usual)
6	Over the past 7 days has your RA stiffness been as variable (coming and going) as usual for you?	5-option Likert scale (It has been much less variable than usual, It has been less variable than usual, It has been the same as usual, It has been more variable than usual, It has been much more variable than usual)
7	Over the past 7 days have you experienced RA stiffness after a period of immobility (for example, after sitting for a while)?	4-option Likert scale (No, not at all, Yes, a little, Yes, quite a lot, Yes, very much)
8	Have you experienced RA stiffness in your body (outside of your joints) over the past 7 days?	4-option Likert scale (No, not in any part of my body, Yes, in a few parts of my body, Yes, in many parts of my body, Yes, all over my body)
9	Has RA stiffness affected your sleep?	4-option Likert scale (Not at all, A little, Quite a lot, Very much)
10	Has RA stiffness made it difficult to dress or undress yourself?	4-option Likert scale (Not at all, A little, Quite a lot, Very much)
11	Has RA stiffness made it difficult to wash yourself (for example, have a shower)?	4-option Likert scale (Not at all, A little, Quite a lot, Very much)
12	Has RA stiffness made it difficult to carry out your responsibilities or commitments?	4-option Likert scale (Not at all, A little, Quite a lot, Very much)
13	Has RA stiffness made it difficult to do your daily tasks or activities?	4-option Likert scale (Not at all, A little, Quite a lot, Very much)
14	Has RA stiffness made it difficult to chew?	4-option Likert scale (Not at all, A little, Quite a lot, Very much)
15	Has RA stiffness made it difficult to do hobbies or activities you enjoy?	4-option Likert scale (Not at all, A little, Quite a lot, Very much)
16	Has RA stiffness made it difficult to get out of bed?	4-option Likert scale (Not at all, A little, Quite a lot, Very much)
17	Has RA stiffness made it difficult to get up after sitting for a while?	4-option Likert scale (Not at all, A little, Quite a lot, Very much)
18	Have your daily tasks and activities required more effort because of RA stiffness?	4-option Likert scale (Not at all, A little, Quite a lot, Very much)
19	Has RA stiffness made you slower (for example, unable to do things quickly)?	4-option Likert scale (Not at all, A little, Quite a lot, Very much)
20	Has RA stiffness made it difficult to do fine movements (for example, write with a pen)?	4-option Likert scale (Not at all, A little, Quite a lot, Very much)
21	Has RA stiffness made it difficult to grip or hold things?	4-option Likert scale (Not at all, A little, Quite a lot, Very much)
22	Has RA stiffness made it difficult to open and close your fist?	4-option Likert scale (Not at all, A little, Quite a lot, Very much)
23	Has RA stiffness reduced your strength to do tasks?	4-option Likert scale (Not at all, A little, Quite a lot, Very much)

Item no.	Item wording	Item response options
24	Has your movement been restricted because of RA stiffness?	4-option Likert scale (Not at all, A little, Quite a lot, Very much)
25	Has RA stiffness made it difficult to balance without physically supporting yourself?	4-option Likert scale (Not at all, A little, Quite a lot, Very much)
26	Have you had to concentrate to move your body because of RA stiffness?	4-option Likert scale (Not at all, A little, Quite a lot, Very much)
27	Have you felt frustrated because of RA stiffness?	4-option Likert scale (Not at all, A little, Quite a lot, Very much)
28	Have you felt worried or concerned because of RA stiffness?	4-option Likert scale (Not at all, A little, Quite a lot, Very much)
29	Have you felt self-conscious because of RA stiffness?	4-option Likert scale (Not at all, A little, Quite a lot, Very much)
30	Has it taken you longer to do your daily tasks or activities because of RA stiffness?	4-option Likert scale (Not at all, A little, Quite a lot, Very much)
31	Have you had to change your plans or behaviour because of RA stiffness?	4-option Likert scale (Not at all, A little, Quite a lot, Very much)
32	Have you had to work around your RA stiffness (or do things in a different way)?	4-option Likert scale (Not at all, A little, Quite a lot, Very much)
33	Have you needed help (from others or gadgets) because of RA stiffness?	4-option Likert scale (Not at all, A little, Quite a lot, Very much)
34	Please circle the number that best describes the impact that RA stiffness has had on your life over the past 7 days	11-point NRS (0=No impact at all, 10=A great deal of impact)
35	Please circle the number that best describes the severity of your RA stiffness over the past 7 days	11-point NRS (0=No stiffness, 10=Extreme stiffness)
36	Please circle the number that best describes how important RA stiffness has been in your life over the past 7 days	11-point NRS (0=Not important at all, 10=Very important)
37	Please circle the number that best describes how well you have coped with your RA stiffness over the past 7 days	11-point NRS (0=Not well at all, 10=Very well)
38	How much of the stiffness you have reported in the questions above is about joints that are permanently stuck?	4-option Likert scale (None of the stiffness I have reported, A little of the stiffness I have reported, Quite a lot of the stiffness I have reported, All of the stiffness I have reported)
39	Please circle the number that best describes the severity of your RA stiffness over a usual week when you are not in a flare?	11-point NRS (0=No stiffness, 10=Extreme stiffness)
40	How would you describe the overall level of morning stiffness you have had from the time you wake up?	11-point NRS (0=No stiffness, 10=Very severe stiffness)
41	How long does your morning stiffness last from waking until maximum improvement occurs?	3-option Likert scale (Up to 1 hour, 1-3 hours, More than 3 hours)
42	Circle the number that best describes the stiffness (all over or in your joints) you felt due to your rheumatoid arthritis during the last week	11-point NRS (0=No stiffness, 10=Very severe stiffness)
43	How would you describe the overall level of morning stiffness you have had from the time you wake up?	5-option Likert scale (No stiffness, Mild stiffness, Moderate stiffness, Severe stiffness, Very severe stiffness)
44	How long does your morning stiffness last from waking until maximum improvement occurs?	Minutes/Hours
45	Were your joints stiff when you woke up today? (Yes/No) If yes, how long did this extra stiffness last?	6-option Likert scale (Less than 30 minutes, 30 minutes to an hour, 1-2 hours, 2-4 hours, More than 4 hours but less than all day, All day)